COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value "Unknown" when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value "Missing" when jurisdictions do not provide a value, and the value "NA" when the value is suppressed as part of privacy protections.

Dataset version: 2021-05-24

Total records in dataset: 25,607,582

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	25,607,582	25,607,582
total_cells	307,290,984	76,822,746
missing_fields	63,082,504	860,778
missing_pct	21%	1%
complete_fields	244,208,480	75,961,968
complete_pct	79%	99%
unknown_fields	21,294,857	9,151,540
unknown_pct	7%	12%
suppressed_fields	99	99
suppressed_pct	0%	0%
available_fields	222,913,524	66,810,329
available_pct	73%	87%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	12	0.0%	632,946	2.5%
sex	17	0.0%	55,183	0.2%
age_group	70	0.0%	172,649	0.7%
records_with_any_field	70	0.0%	800,435	3.1%